ISSN 0956-8549

# Delegated Blocks

**By**

**Amil Dasgupta**
**Richmond Mathews**

**DISCUSSION PAPER NO 866**

**PAUL WOOLLEY CENTRE WORKING PAPER No 91**

**March 2023**

# Delegated Blocks*

Amil Dasgupta† Richmond Mathews‡

This version: February 2023

**Abstract**

Will asset managers with large amounts of capital and high risk-bearing capacity hold large blocks and monitor aggressively? Both block size and monitoring intensity are governed by the contractual incentives of institutional investors, which themselves are endogenous. We show that when high risk-bearing capacity arises via optimal delegation, funds holds smaller blocks and monitor significantly less than proprietary investors with identical risk-bearing capacity. This is because the optimal contract enables the separation of risk sharing and monitoring incentives. Our findings rationalize characteristics of real world asset managers and imply that block sizes will be a poor predictor of monitoring intensity.

*We are grateful to Mark Loewenstein, John Moore, Shri Santosh, Dimitri Vayanos and seminar audiences at the LSE and the University of Maryland at College Park for helpful comments. Dasgupta acknowledges financial support from the ESRC via Research Grant ES/S016686/1.

†London School of Economics and ECGI. Email: a.dasgupta@lse.ac.uk

‡Smith School of Business, University of Maryland at College Park. Email: rmathews@umd.edu

# 1 Introduction

The rise of asset managers has led to the concentration of vast amounts of capital in the hands of institutional investors.[1] How is this likely to affect corporate governance? These investors have large risk-bearing capacity and thus are able—in principle—to hold large blocks and monitor portfolio firms aggressively. However, both block size and the extent of monitoring are endogenous to the contractual incentives of institutional investors. Such contractual incentives—in turn—are endogenously determined, and will anticipate institutional ownership and monitoring decisions. Will institutional investors hold large blocks commensurate to their risk-bearing capacity in equilibrium? Conditional on holding such blocks, will they monitor firms aggressively? To help answer these questions, we study the economics of delegated blockholding. In particular, we characterize corporate governance and risk sharing in markets where equity ownership is optimally delegated and both equity block sizes and the level of monitoring are determined by endogenous contracts established between asset managers and their investors.

We benchmark our analysis against the influential characterization of risk sharing and monitoring in a market with *proprietary* ownership found in Admati, Pfleiderer, and Zechner (1994)—APZ henceforth. Taking as given the existence of a proprietary trader with high risk-bearing capacity, APZ consider whether anticipated monitoring costs will limit the trader's willingness to hold large blocks. Under broad and plausible conditions, they find the answer is "no"—as long as traders with high risk-bearing capacity cannot commit to limit their trading, they will trade to the competitive risk sharing allocation and monitor at a level consistent with that allocation. This is because the ability to trade repeatedly erodes the large traders' strategic advantage. APZ's striking finding is confirmed in the fully dynamic analysis of DeMarzo and Urosevic (2006). Overall, therefore, the existing

[1] See, e.g., Dasgupta, Fos, and Sautner (2021) for relevant stylized facts.

literature provides a reassuring view: risk sharing and monitoring can coexist happily in financial markets as long as ownership is proprietary.

We show that when high risk-bearing capacity is instead attained endogenously via delegation, outcomes are dramatically different. First, the optimal fund holds less of the risky asset, i.e., a smaller block, than an investor with the same risk-bearing capability would under the competitive risk sharing allocation. In other words, delegation hurts risk sharing. Second, delegation separates block sizes and monitoring incentives, because monitoring is undertaken by professional asset managers on behalf of the fund. It is *their* effective stake, not the fund's overall stake, that determines the fund's level of monitoring. The optimal delegation contract allocates an effective stake to these professional asset managers that results in a level of monitoring that would be privately optimal for fund investors at their *initial* endowment. These two effects combined imply that the optimal fund undertakes significantly less monitoring than a proprietary blockholder of identical risk-bearing capacity. While delegation thus has negative implications for risk sharing and monitoring relative to the case with proprietary large traders, it does provide valuable risk sharing opportunities to agents who do not have full access to financial markets.

**Model summary.** We start with a minor variation of the APZ benchmark. Our version of their classical "CARA-Normal" model features a firm whose final-date equity cash flows are distributed Normally and a group of traders with CARA utility whose aggregate risk tolerance (i.e., risk-bearing capacity) sum to unity. The traders are made up of a single large trader $L$ with risk tolerance of $\lambda$ and a continuum of small traders whose aggregate risk tolerance is given by $1 - \lambda$. In addition to trading (potentially many times) in a Walrasian market at the initial date, $L$ can also monitor at an intermediate date: such monitoring is costly for $L$ but increases average final-date cash flows to all equity holders. The competitive equilibrium allocation in such an economy involves $L$ holding $\lambda$ fraction of the firm's equity.

Imagine that $L$ starts with an endowment $\omega < \lambda$. Will $L$ trade from $\omega$ all the way to $\lambda$? There are several impediments. First, $L$ knows that if she trades to $\lambda$ she will then monitor at a commensurately higher intensity and all $1 - \lambda$ other shareholders will benefit from such monitoring. Second, $L$ knows that along the way to $\lambda$ she must pay the full value of future monitoring when acquiring shares, i.e., she moves prices against herself as she trades. However, in a key result, APZ show that as long as $L$ can't commit to limit her trading, she will nevertheless trade to $\lambda$ and monitor at the high intensity corresponding to such large holdings. This arises because of an endowment effect. Counterfactually, if any sequence of trades led to a proposed final holding level for $L$ that is strictly below $\lambda$, she would be tempted to buy a bit more because the current holding is now part of her endowment. Starting from this endowment, there will always be at least some incremental risk-sharing gains by buying a bit more, despite having to pay the full value of future monitoring in making such purchases. This result implies that the anticipation of future monitoring costs does not act as an impediment to holding large positions in financial markets.

In our main analysis, we enrich the APZ framework to model delegated blockholding. We replace $L$ with a measure of small, constrained, investors with aggregate risk tolerance $\lambda$ who hold an aggregate endowment of $\omega < \lambda$ (the same as $L$ above) but cannot trade directly. We allow the continuum of unconstrained investors with aggregate risk tolerance $1 - \lambda$ to offer trading and monitoring services to the constrained investors in return for fees. A fund is formed if constrained investors (who we then call limited partners, or LPs) team up with an endogenously chosen subset of unconstrained investors with aggregate risk tolerance $\tau \in (0, 1 - \lambda]$ (who we then call general partners, or GPs) to pool their endowments. LPs are passive once the fund is formed, and GPs determine holding and monitoring levels subject to an contract specifying a fee $f$ paid by LPs to join the fund and a skin in the game parameter $\phi \in [0, 1]$ representing the GPs' share of the fund's assets. Since GPs can choose to unilaterally deviate from the fund and benefit from the monitoring undertaken by the fund, the contractual payments must compensate GPs for their monitoring costs. Subject to compensating the GPs for their costs, the contract aims to induce them to trade and monitor in a way that is desirable for LPs. We solve for the optimal linear contract by choosing the fund parameters $(\tau^*, \phi^*, f^*)$ that maximize LP payoffs subject to the GPs participation constraint, while ensuring that no LP would unilaterally withhold their endowment from the fund. Our optimal linear contract also turns out to be the overall optimal contract.

We show that—despite the fact that the fund cannot commit to limit its trade (exactly as in APZ)—the optimal contract induces radically different trading and monitoring choices relative to the APZ benchmark. A key insight is that delegation separates monitoring incentives from overall holdings. This is because delegated monitoring is undertaken by professional asset managers on behalf of the fund: It is *their* effective stake, not the fund's overall stake, that determines the fund's level of monitoring. The optimal contract allocates a share of the fund's assets to GPs that induces monitoring at a level consistent with only the LPs' *initial* endowment; in other words, LPs do not have to compensate GPs for any monitoring that is excessive from the LP's private perspective. However, since LPs' initial endowment is $\omega < \lambda$, whereas the aggregate risk-bearing capacity of the LP's is $\lambda$, the optimal fund monitors less than a proprietary trader with identical risk-bearing capacity. Further, we show that the fund also holds too small an overall position in the asset: in particular, under the optimal contract, the LPs hold a position within the fund that fully reflects their market power as a strategic trader with aggregate risk tolerance $\lambda$. Overall, therefore, by separating monitoring incentives from risk sharing, the optimal contract enables LPs to attain their *privately* optimal, full-commitment, levels of both monitoring and risk sharing. But this is attained at the expense of lower overall levels of monitoring and risk sharing in the market. That said, the ability to access financial markets via delegation clearly enhances risk sharing relative to the case where constrained agents are simply excluded from financial markets.

**Applied implications.** Our main results characterize the economics of monitoring and risk sharing in financial markets with delegated blockholding. Given the preponderance of delegated asset managers in modern financial markets, these results are relevant to interpreting key features of blockholding and monitoring that are prevalent today. Specifically, our model has three main applied implications for corporate governance and the role of the asset management industry.

*Which asset managers will monitor.* Our analysis of optimal delegation arrangements has implications for the degree to which different types of asset managers should be expected to engage in the monitoring of portfolio firms. In particular, we show that asset managers' (i.e., GPs') skin in the game, which determines their level of monitoring, is increasing in the endowment of each underlying investor (LP) in the fund. Thus, if fund investors have relatively high endowments, they will invest in funds in which managers take larger personal stakes and monitor aggressively. If, on the other hand, fund investors have relatively low endowments, they will invest in funds in which managers will take small personal stakes and monitor very little.

This depiction resonates with key characteristics of asset management firms observed in reality. Relatively poor real-world investors tend to invest in mutual funds. It is well documented that mutual fund managers have very little self-investment in their funds (Khorana, Servaes, and Wedge 2007), and mutual funds are notorious for being muted in their engagement efforts (e.g., Bebchuck et al 2017). In contrast, wealthy individuals tend to invest in hedge funds. Managers of these funds are well known to self-invest significantly and play an active role in the monitoring of their portfolio firms (Agarwal, Daniel, and Naik, 2009, Brav, Jiang, and Kim 2010).

*Larger blocks may monitor less than small blocks.* Our results imply that block size may not be a good predictor of monitoring intensity. With proprietary blocks as in APZ, larger stakes imply more monitoring because stake size directly determines monitoring intensity. However, with delegated blocks the fund's internal incentive structure separates monitoring incentives from stake size. In particular, in our model, the endogenous block size is increasing in both the number of fund investors and their initial endowments, whereas monitoring intensity is determined only by their initial endowment. As a result, blocks held by funds with many investors with low initial endowment may be larger but feature significantly less monitoring than those held by funds with a smaller number of investors with higher initial endowment. In this regard, our results are consistent with Nockher (2022), who shows that smaller blockholders tend to be more intensive monitors than larger blockholders.

*The role of index funds in governance.* At the broadest level, our analysis indirectly highlights a role for index funds in corporate governance. Our analysis speaks to the concentrated holding choices of active funds, who make deliberate portfolio decisions (as our GPs do). A key implication of our model is that such active asset managers do not utilize their full risk-bearing capacity to hold concentrated positions, and expend suboptimally low levels of resources into monitoring. This finding must be viewed in the context of the evolution of the asset management industry and the emergence of passive, i.e., index, funds which—purely by virtue of their size—mechanically end up holding concentrated positions in firms. If active funds do not hold sufficiently concentrated stakes and thus limit their monitoring, as our results suggest, it becomes all the more important to understand the role of index funds in governance (Brav, Malenko, and Malenko 2022).

### 1.1 Related literature

Our paper builds most directly on APZ and papers that generalize the model of APZ, e.g., DeMarzo and Urosevic (2006), who confirm APZ's key conclusion in a fully dynamic model. However, our work relates more broadly to a number of literatures.

At the most basic level, our paper is connected to the significant theoretical literature that studies blockholder monitoring. This literature is surveyed by Edmans and Holderness (2017). Many papers within this literature (e.g., Shleifer and Vishny 1986, Faure-Grimaud and Gromb 2004) take block size as being exogenous. Others (e.g., Kyle and Vila 1991, Maug 1998, Kahn and Winton 1998, Back, Collin-Dufresne, Fos, Li, and Ljunqvist 2018) consider how proprietary blocks can emerge endogenously by focusing on the ability to generate short-term trading profits. Our analysis differs from all these prior papers by explicitly modeling the emergence of *delegated* equity blocks. Further, in contrast to the second strand discussed above, we assume fully transparent financial markets, so there are no trading profits; in this respect, our analysis has similarities to Bolton and von Thadden (1998), though they also focus purely on proprietary blocks.

More recently, a growing theoretical literature takes the delegated nature of equity ownership seriously, and considers the role of the incentives of asset managers in corporate governance. This literature is surveyed by Dasgupta, Fos, and Sautner (2021) (see, in particular, section 4 of that paper). While several papers within that literature (e.g., Dasgupta and Piacentino 2015) have highlighted the negative implications of agency frictions arising from the delegation of portfolio management on the level of monitoring at portfolio firms, none of those papers endogenize the presence of delegated blockholders.

Finally, our paper is related in spirit to the literature on the endogenous emergence of financial intermediaries, starting with the work of Diamond and Dybvig (1983), as well as the literature on optimal contracting in delegated portfolio management, starting with the work of Bhattacharya and Pfleiderer (1985). Relative to the former, which has focussed on banking, we consider the emergence of asset managers. Relative to the latter, which considers optimal contracting with respect to trading by asset managers, we incorporate monitoring considerations as well.

## 2 A benchmark model

We start with a simplified, benchmark, version of the APZ model. Consider a financial market with a single firm with 1 infinitely divisible equity share outstanding, and a risk-free asset in perfectly elastic supply whose gross return is normalized to unity. There is a unit continuum of traders who have CARA utility, each with risk tolerance of 1. To mirror the assumption of an exogenously specified large trader in APZ, we assume that a measure $\lambda$ of such traders are aggregated into a single strategic trading entity, $L$, who trades strategically taking her price impact into account, and can monitor the firm to improve its cash flows. The remaining $1-\lambda$ of atomistic traders trade competitively and cannot monitor. We assume that $L$ has an endowment of $\omega \in (0, \lambda]$ shares while the remaining $1-\lambda$ traders have an aggregate endowment of $1-\omega$ shares, shared equally among them.

There are three dates. Potentially numerous rounds of trading opportunities are available at date 1 in a Walrasian market: in any given round of trade, traders submit demand functions and a market-clearing price is determined. At date 2, $L$ can choose to monitor the firm as follows: at a cost of $\frac{1}{2}\gamma m^2$, where $\gamma > 0$, she can exert monitoring effort $m \geq 0$ to generate a final equity payoff that is distributed according to $N\left(\mu + m, \sigma^2\right)$. At date 3, all payoffs are publicly realized. As in the bulk of the APZ analysis, $L$ cannot commit to a final round of trade at date 1 or to a particular level of monitoring at date 2.[2]

**Aggregate risk tolerance.** In a Walrasian CARA-Normal market with symmetric information like ours, each competitive agent with unit risk tolerance will have a demand function of $\frac{m+\mu-P}{\sigma^2}$, and thus the total demand of a measure $x$ of atomistic competitive agents is given by $x\frac{m+\mu-P}{\sigma^2}$, which is equivalent to the demand of a *single* competitive agent with risk tolerance of $x$. In other words, with unit underlying risk tolerance, the *aggregate risk tolerance* of a given measure of atomistic competitive agents is given by the measure of those agents. Accordingly, throughout the paper, we shall treat the $1-\lambda$ of atomistic traders as being represented by a single competitive trader with risk tolerance of $1-\lambda$. For benchmarking purposes, we assume that $L$ has the same risk tolerance as the aggregate risk tolerance of the measure of competitive agents he replaces, i.e., $\lambda$. This assumption will be convenient when we generalize the model to explicitly model the large trader as an endogenous delegated trading vehicle, i.e., a fund, formed of a measure of investors and fund managers.

**Competitive allocations with perfect risk sharing.** Before analyzing the full equilibrium involving both strategic and competitive trading as well as monitoring, it is helpful to establish a benchmark in which all traders are competitive and monitoring cannot arise. In such a benchmark, risk sharing considerations are the sole determinants of equilibrium allocations. Denoting $L's$ equilibrium holdings by $\alpha$, it is easy to see that the *competitive equilibrium allocation* is $\alpha = \lambda$. This is because the competitive equilibrium involves perfect risk sharing, under which $L$ would hold $\frac{\lambda}{\lambda+1-\lambda} = \lambda$ fraction of the risky asset while the atomistic traders would hold $\frac{1-\lambda}{\lambda+1-\lambda} = 1-\lambda$ of the risky asset, in accordance with their

[2]APZ also consider the case of multiple assets as well as more general monitoring technologies; we use this baseline version of their model, as it is under these specific assumptions that APZ provide the most complete characterizations.

relative levels of risk tolerance.

**Equilibrium trading and monitoring.** In order to analyze $L's$ trading, taking into account both strategic and monitoring incentives, we follow APZ to outline a few baseline steps. First, given that $L$ is unable to commit to a particular level of monitoring, the equilibrium monitoring level is determined by L's final holdings on date 2. If $\alpha$ is L's total ownership of the risky asset upon entering date 2, then $m$ is given by:

$$m(\alpha) = argmax_m \alpha(\mu + m) - \frac{1}{2}\gamma m^2 = \frac{\alpha}{\gamma}. \tag{1}$$

Note that $m$ does not affect the risk of $L$'s portfolio, so risk adjustment does not affect this choice.

If the $1 - \lambda$ atomistic investors anticipate that $L's$ final ownership of the risky asset is given by $\alpha$, they have an aggregate demand of

$$(1 - \lambda)\,\frac{\mu + m(\alpha) - P}{\sigma^2},$$

giving rise to a market clearing price of

$$P\left(\alpha\right) = \mu + m(\alpha) - \frac{1 - \alpha}{1 - \lambda}\sigma^2. \tag{2}$$

Finally, given that $L$ is unable to commit to a final round of trade within date 1, we follow APZ in focusing on *globally stable allocations.* In the absence of the ability to commit to a given number of trades, APZ show that following any sequence of trades, $L$ will wish to trade again ahead of her monitoring choice unless she has traded to an allocation which is globally stable. Such an allocation is defined as follows:

**Definition 1.** An allocation $\alpha_G$ is globally stable iff (i)

$$\alpha_G \in argmax_\alpha \Psi(\alpha) - \Psi(\alpha_G) - (\alpha - \alpha_G)P(\alpha_G),$$

and (ii) for every $\omega \in [0,1]$, such that $\omega \neq \alpha_G$,

$$\Psi(\alpha_G) - \Psi(\omega) - (\alpha_G - \omega)P(\alpha_G) > 0,$$

where

$$\Psi(\alpha) = \alpha(\mu + m(\alpha)) - \frac{1}{2}\gamma m(\alpha)^2 - \frac{1}{2\lambda}\alpha^2\sigma^2 \tag{3}$$

is the certainty equivalent for $L$ of holding $\alpha$ units of the risky asset and monitoring accordingly.

In words, this means that: (i) once a globally stable allocation is reached, $L$ will not wish to trade away from it at current prices; and (ii) $L$ is willing to trade to the globally stable allocation from any other position at prices consistent with the globally stable allocation. In their central result, APZ show that:

**Proposition 1. (Admati, Pfleiderer, and Zechner 1994)** *As long as $\lambda < \gamma\sigma^2$,there exists a unique globally stable allocation, $\alpha_G = \lambda$, which coincides with the competitive equilibrium allocation.*

All proofs are in the Appendix. The restriction $\lambda < \gamma\sigma^2$ ensures the concavity of the large trader's certainty equivalent (3). It is a special case of a more general concavity assumption in APZ (see APZ's Proposition 3 and 4).

This key result implies that the possibility of monitoring does not affect the degree of risk sharing in equilibrium. The reason, as APZ discuss, is that the lack of the ability to commit to a final round of trade erodes the strategic advantage of the strategic trader, who subsequently trades all the way to the competitive equilibrium allocation. Put another way, there is no trade-off between diversification and monitoring because an endowment effect induces the large trader, $L$, to trade all the way to the risk-sharing optimum. Counterfactually, if any sequence of trades led to a proposed final holding level for $L$ that is strictly below her risk-sharing optimum, she would be tempted to buy a bit more because the current holding is now part of her endowment. Starting from this endowment, there will always be at least some incremental risk-sharing gains by buying a bit more, despite having to pay the full value of future monitoring in making such purchases.

While the concept of global stability is essentially static, its relevance has been confirmed by DeMarzo and Urosevic (2006) in a fully dynamic version of the APZ model with constant trading and monitoring opportunities. Indeed, their main result is that the large trader will ultimately trade to the competitive price-taking allocation, which generalizes and provides a dynamic micro-foundation for APZ's concept of global stability.

In reality, the concentration of ownership into the hands of a large trader is typically achieved by delegating portfolio management to professional asset managers who trade and monitor on behalf of their clients. Thus, in the remainder of the paper, we examine how incentives to monitor are determined when risk averse investors can optimally delegate to a professional fund manager, who can then trade freely in financial markets but cannot make prior commitments to monitor firms at any particular level of intensity.

## 3 Delegated blocks

We now introduce the possibility of delegated blockholding. Instead of assuming that a measure $\lambda$ of investors exogenously acts as a single trading entity as above, we now assume that a measure $\lambda$ of investors lacks the ability to trade directly in the market and are thus "constrained." If such investors wish to trade the risky asset, they must pool their resources and employ a professional asset manager, thus endogenously generating the possibility of delegated blockholding. The remaining $1 - \lambda$ fraction of "unconstrained" investors can trade freely in markets, as in the benchmark model. Such agents also have the ability to group themselves into "funds" to offer investment services to the constrained investors. To be consistent with the benchmark model, the $\lambda$ measure of constrained investors has aggregate endowment $\omega \leq \lambda$ of the risky asset (shared equally), while the $1 - \lambda$ measure of unconstrained investors have the remaining $1 - \omega$ endowment (also shared equally, as in the benchmark).

A fund is formed when the $\lambda$ measure of constrained investors, who we also refer to as Limited Partners, or "LPs," decide to employ a chosen measure $\tau \in (0, 1 - \lambda]$ of unconstrained investors, who we also refer to as General Partners, or "GPs," to trade and monitor on their behalf.[3] The GPs in a fund act collectively to make trading and monitoring decisions based on their joint incentives, while the LPs are passive once they have joined the fund. As in the benchmark model, the GPs cannot commit to a given trading strategy or monitoring level up front—they always behave opportunistically once the fund has been established. When a fund is formed, all GPs and LPs joining the fund contribute their endowments to the fund and agree on fees and ownership stakes, which we restrict to be linear as follows: a skin in the game parameter, $\phi \in [0, 1]$, specifying the GPs' share of the fund's assets, and an up-front fee, $f$, which each participating LP must pay to join the fund. Overall, a fund can thus be represented as a linear contracting triple, $(\tau, \phi, f)$, representing the measure of GPs, their skin in the game, and the per-LP

[3] Note that no funds would form without the participation of constrained investors. In any fund with only unconstrained investors, some subset of those investors must monitor and thus pay costs. However, any investor that is supposed to be in the subset that monitors can choose not to join the fund, trade on their own, and enjoy exactly the same cash flow payoffs without paying the monitoring costs. So, to persuade them to join the fund, the subset that are in the fund but do not monitor must pay those that are expected to monitor. But this effectively means that monitoring costs are shared among all agents in the fund, and so the previous argument applies and individual investors who are not expected to monitor would prefer not to join the fund.

fee, respectively.

Since our interest is in optimal delegation, we aim to maximize the payoff of LPs. In other words, the linear contracting terms $(\tau, \phi, f)$ are chosen to optimize the payoff of the LPs while ensuring the participation of the requisite mass of GPs. Our choice can be microfounded in terms of perfect competition among potential GP groups for the right to form the fund with the LPs. Indeed, we show later (after Proposition 2) that as long as the proportion of constrained agents in the population is not too large, we can microfound our analysis in terms of Bertrand competition among groups of GPs. Finally, while we undertake our analysis within this linear space of contracts, we show below that the optimal linear contract is also an optimal contract (Corollary 2).

We solve the model by backward induction. We first assume that a fund with $\lambda$ LPs and (an arbitrary positive measure of) $\tau$ GPs is formed, and proceed to compute the monitoring and trading decisions of the GPs for a given $(\tau, \phi, f)$. We then solve for the optimal linear contracting terms. To do this, we account for the fact that, for each measure of GPs $\tau$, the fees $\phi$ and $f$ will be determined endogenously. We then find the optimal $\tau$, ensuring that all $\tau$ GPs are willing to join the fund. We denote the optimal set of linear contracting terms by the triple $(\tau^*, \phi^*, f^*)$. Finally, we check that it is incentive compatible for all $\lambda$ LPs to join the fund under these terms.

We reuse $\alpha$ to denote the final stake in the risky asset held by the fund. The GPs have an *effective stake* in the final payoff of the risky asset equal to their proportional share of the fund's stake, or $\phi\alpha$. Given that monitoring does not affect the risk of their payoff, they will optimally choose $m$ as follows (all $D$-superscripts refer to functions defined for the delegated fund model):

$$m^D(\alpha) = argmax_m \phi\alpha(\mu + m) - \frac{1}{2}\gamma m^2 = \frac{\phi\alpha}{\gamma}. \tag{4}$$

Since the GPs cannot commit to a given trading strategy, we again focus on globally stable trading allocations. We note first that the pricing function must be adjusted for the fact that the mass of competitive price-taking investors has been reduced from $1-\lambda$ to $1-\lambda-\tau$ given the formation of the fund. If the competitive investors expect the fund to end up with a stake of $\alpha$, their aggregate demand will be

$$(1-\lambda-\tau)\,\frac{\mu+m^D(\alpha)-P}{\sigma^2},$$

giving rise to a market clearing price of

$$P^D(\alpha)=\mu+m^D(\alpha)-\frac{1-\alpha}{1-\lambda-\tau}\sigma^2. \tag{5}$$

The definition of a globally stable allocation must also be adjusted for our delegated fund model as follows, since GPs make decisions on behalf of the entire fund but enjoy only a $\phi$ proportion of its payoff.

**Definition 2.** An allocation $\alpha_G^D$ is globally stable iff (i)

$$\alpha_G^D \in argmax_\alpha \Psi^D(\alpha)-\Psi^D(\alpha_G^D)-\phi(\alpha-\alpha_G^D)P^D(\alpha_G^D),$$

and (ii) for every $\omega\in[0,1]$, such that $\omega\neq\alpha_G^D$,

$$\Psi^D(\alpha_G^D)-\Psi^D(\omega)-\phi(\alpha_G^D-\omega)P^D(\alpha_G^D)>0,$$

where

$$\Psi^D(\alpha)=\phi\alpha(\mu+m^D(\alpha))-\frac{1}{2}\gamma m^D(\alpha)^2-\frac{1}{2\tau}\phi^2\alpha^2\sigma^2 \tag{6}$$

is the certainty equivalent for the GPs if the fund holds a stake of $\alpha$ units of the risky asset and they monitor accordingly. We have the following result.

**Lemma 1.** *As long as $\tau < \gamma\sigma^2$, there exists a unique globally stable allocation*

$$\alpha_G^D = \frac{\tau/\phi}{\tau/\phi + 1 - \lambda - \tau}. \tag{7}$$

The equilibrium stake of the fund $\alpha_G^D$ depends only on the sizes of the LP and GP populations, $\lambda$ and $\tau$, and the skin in the game parameter, $\phi$. The expression is analogous to the globally stable allocation in the APZ benchmark derived in Proposition 1: it is part of the competitive equilibrium allocation in a market with two competitive traders, one of whom has risk tolerance of $1 - \lambda - \tau$ while the other has risk tolerance of $\tau/\phi$. The former is simply the aggregation of the unconstrained investors who did not join the fund (i.e., did not become GPs). The latter investor aggregates the unconstrained investors inside the fund, i.e., the GPs. Recall that trading decisions in the fund are taken by a measure $\tau$ of GPs who have an aggregate risk tolerance of $\tau$. However, these GPs are only exposed to a fraction $\phi$ of the holdings of the fund, giving them an *effective risk tolerance* of $\tau/\phi$.

Notably, however, this allocation does not necessarily correspond to perfect risk sharing among all investors—which arose in the globally stable allocation of APZ (see Proposition 1)—as this would require an allocation of $\alpha_G^D = \tau + \lambda$ (since the measure of investors in the fund is the sum of $\lambda$ LPs and $\tau$ GPs). The deviation from perfect risk sharing arises due a combination of two factors: First, only a measure of $\tau < \tau + \lambda$ agents make decisions on behalf of the whole fund; and second, those agents are exposed to only a fraction $\phi$ of the fund's holdings. Indeed, it is apparent that if $\tau/\phi = \tau + \lambda$ in the expression for $\alpha_G^D$ above, we obtain $\alpha_G^D = \tau + \lambda$. Given this deviation from perfect risk sharing in equilibrium, the determination of the optimal linear contracting parameters $(\tau^*, \phi^*, f^*)$ is critical for determining both the level of monitoring and the degree of diversification in the model.

The restriction $\tau < \gamma\sigma^2$ is analogous to the condition $\lambda < \gamma\sigma^2$ in Proposition 1 and ensures that $\Psi^D(\alpha)$ is strictly concave. We impose this restriction throughout the remainder of the analysis.

Comparing the effective stakes of unconstrained investors inside and outside the fund yields the following result.

**Lemma 2.** *The GPs in the fund and the outside unconstrained investors end up with identical effective per-investor holdings of the risky asset.*

This result implies that there is perfect risk sharing over the part of the risky asset that is not (effectively) held by the LPs among the total $1 - \lambda$ measure of unconstrained investors, whether inside or outside the fund. This is because the existence of multiple trading opportunities combined with the inability to commit to a particular trading strategy erodes the strategic advantage of the GPs, who subsequently trade to arrive at the point of perfect risk sharing between themselves (with effective risk tolerance $\tau/\phi$) and unconstrained investors outside the fund (with aggregated risk tolerance $1 - \lambda - \tau$), as discussed above.

Given the trading and implied monitoring choices of the GPs for a given $(\tau, \phi, f)$, we now proceed to endogenize the linear contracting terms to determine $(\tau^*, \phi^*, f^*)$. In our central result, we show that:

**Proposition 2.** *There exists a $\hat{\omega} \in (0, \lambda)$ such that for $\omega \leq \hat{\omega}$, delegated blockholding arises in equilibrium and the optimal linear delegation parameters are characterized by:*

1. *a mass of GPs $\tau^* = \frac{(1-\lambda^2)\omega}{1-\lambda\omega}$,*

2. *a skin in the game parameter $\phi^* = \frac{(1+\lambda)\omega}{2\lambda\omega+\lambda+\omega}$, and*

*3. a fee*

$$f^* = \frac{1}{\lambda}\left(\begin{array}{c} \frac{\tau^*}{1-\lambda-\tau^*}\left[\Psi_U^D(\alpha_G^D) - \left(1-\alpha_G^D - \left(1-\omega-\tau^*\frac{(1-\omega)}{1-\lambda}\right)\right)P^D(\alpha_G^D)\right] \\ -\left[\Psi^D(\alpha_G^D) - \phi(\alpha_G - \omega - \tau^*\frac{(1-\omega)}{1-\lambda})P^D(\alpha_G^D)\right] \end{array}\right) \quad (8)$$

*where* $\alpha_G^D$, $\Psi^D$, *and* $\Psi_U^D(\alpha) = (1-\alpha)(\mu_0 + m^D(\alpha)) - \frac{(1-\alpha)^2\sigma^2}{2(1-\lambda-\tau)}$ *are each evaluated at* $\phi^*$ *and* $\tau^*$.

*If* $\omega > \hat{\omega}$*, delegated blockholding does not arise in equilibrium.*

The proof proceeds in several steps. First, we fix $\tau$ and $\phi$ and obtain the fee $f$ as a function of these two parameters. Since our goal is to maximize LP payoffs, the fee $f$ is set to just satisfy the participation constraint of individual GPs conditional on the existence of a fund involving $\tau$ GPs with a skin in the game parameter $\phi$. Given$f(\tau, \phi)$ and for any given $\tau$, we then maximize the per LP payoff over $\phi$ to obtain the optimal GP share $\phi(\tau)$ for each $\tau$. Next, we maximize the induced per-LP payoff to obtain the optimal $\tau$, $\tau^*$, which then gives us the remaining optimal parameters $\phi^* \equiv \phi(\tau^*)$ and $f^* \equiv f(\phi^*, \tau^*)$ shown in the proposition. Finally, we show that, as long as individual LPs' endowments are not too high ($\omega \leq \hat{\omega}$), LPs are willing to join the optimal fund rather than stay out of the fund and consume their endowment. We next examine the implications of this result and the intuition behind it through a series of remarks and corollaries.

First, we examine the role of the fee, $f^*$. Any individual unconstrained agent can choose between joining the measure $\tau$ of GPs inside the fund or remaining part of the measure $1-\lambda-\tau$ of unconstrained agents who do not join the fund. The fee $f^*$ is set to make the payoffs of these two choices equal.

*Remark* 1. The fee $f^*$ compensates GPs for both their expected equilibrium monitoring costs as well as excess trading costs they incur relative to outside unconstrained investors.

Since, as shown in Lemma 2, GPs in the fund and unconstrained investors outside the fund end up holding the same effective stake per investor, their degree of diversification is not affected by joining the fund and the LPs do not need to compensate them for taking more or less risk. However, unconstrained agents choosing to become GPs inside the fund share the cost of monitoring, while those remaining outside the fund enjoy the benefits of such monitoring for free. Thus, the fee $f(\tau, \phi)$ must compensate GPs for their monitoring costs at the fund's ultimate stake. In addition, the fee must compensate GPs for any excess trading cost (relative to outside unconstrained investors) in arriving at their final post-trade allocation from their initial contractually-induced endowment. The reason that the GPs incur such excess costs is that their initial pre-trade endowment under the contract, $\frac{\phi^*}{\tau^*}\left(\omega + \tau^* \frac{1-\omega}{1-\lambda}\right)$, is actually smaller than $\frac{1-\omega}{1-\lambda}$, the initial endowment of each unconstrained agent. Thus, in order to optimally share risk with outside unconstrained agents, GPs must trade more than their counterparts outside the fund.

Next, we note from the proposition that the maximal endowment per constrained investor consistent with the existence of delegated blockholding, $\hat{\omega}$, is strictly smaller than their measure in the population of agents, i.e., $\hat{\omega} < \lambda$. Thus, each constrained agent's endowment of the risky asset must be strictly smaller than that of each unconstrained agent: $\frac{\hat{\omega}}{\lambda} < 1 < \frac{1-\hat{\omega}}{1-\lambda}$. In other words:

*Remark* 2. Delegated blockholding can arise only when constrained agents have relatively low endowments of the risky asset.

In order to derive intuition for why delegated blockholding cannot arise when constrained agents have higher endowments, we establish a key intermediate result. Utilizing the expressions for the fund's overall equilibrium stake $\alpha_G^D$ from Lemma 1 and the optimal skin in the game parameter $\phi^*$ and optimal GP pool $\tau^*$ from Proposition 2, we can compute the effective stakes of LPs ($(1-\phi^*)\,\alpha_G^D$) and GPs ($\phi^* \alpha_G^D$).

**Corollary 1.** *The effective aggregate stake of GPs in the fund is $\omega$, while the effective aggregate stake of LPs is $\lambda\frac{1+\omega}{1+\lambda}$.*

We begin with a simple observation about the effective aggregate stake of the LPs under optimal delegation: $\lambda\frac{1+\omega}{1+\lambda}$. Since $\omega < \lambda$, the effective aggregate stake of the LPs lies between their initial endowment ($\omega$) and their unconstrained risk-sharing holding level ($\lambda$):

$$\omega < \lambda\frac{1+\omega}{1+\lambda} < \lambda.$$

In other words, under optimal delegation, constrained LPs are taken towards their unconstrained risk-sharing optimal holding level but not all the way: delegation aids risk sharing for constrained LPs. This provides some intuition for why delegation can only be successful when constrained LPs have relatively low endowments in the risky asset (Remark 2). If the endowment $\omega$ was sufficiently close to the unconstrained risk sharing level, then risk-sharing gains would be small, and individual LPs could simply defect from the fund, avoiding the payment of fees, but enjoying any monitoring that would be undertaken by the fund anyway in their absence.

We now turn to a broader discussion of the levels of risk-sharing and monitoring implied by Corollary 1. The optimal delegation contract selects the skin in the game parameter ($\phi^*$) and measure of GPs ($\tau^*$) so as to render the effective holdings of the GPs to be equal to the endowment of the LPs ($\omega$). This means that GPs (who undertake monitoring) will monitor "as if" they hold the original endowment of the LPs. LPs, however, do *not* hold their original endowment, but rather end up with a larger holding of the risky asset: $\lambda\frac{1+\omega}{1+\lambda} > \omega$. To obtain intuition for these two facts, it is helpful to separately consider the monitoring and risk-sharing incentives of LPs.

Let us begin by shutting down risk-sharing incentives. Imagine, counterfactually, that the LPs are exogenously aggregated into a single *risk-neutral* collective with endowment $\omega$, whereas all other agents are atomistic and do not monitor. In this case, the LPs would be the sole monitor, paying the full cost of monitoring, but enjoying only an $\omega$-portion of the benefits. Furthermore, since there are no risk sharing considerations, they would have no incentive to purchase more shares because that would induce them to monitor more, while at the same time they would have to pay out the full benefits of the increased monitoring to purchase those shares. This is due to the well-known free-riding phenomenon (first formalized by Grossman and Hart 1980). So effectively, the LPs couldn't "enjoy" the monitoring benefits on anything more than their initial endowment. Thus, absent any diversification incentives, LPs would simply hold their initial endowments and monitor at a level consistent with such endowments.

Of course, in the delegated case, LPs do not directly monitor, but GPs monitor on their behalf. However, the LPs must pay the *full* cost of monitoring via the fees paid to the GPs. Why is this, when the GPs would also benefit from the monitoring? The answer lies in the GPs' outside option: to defect from the fund and trade on their own account while enjoying the monitoring by the remaining GPs. The fees must be set to prevent such a defection, and thus encumber the LPs with the full monitoring cost to keep the GPs in the fund. Furthermore, any shares purchased by the fund will, like shares purchased by a proprietary blockholder, be purchased at a price that fully reflects any additional monitoring. As a result, the LPs would find it optimal, *absent risk sharing motives*, to constrain equilibrium monitoring to the level of their initial endowment $\omega$. Of course, LPs are not risk-neutral in our model, and the question then arises immediately: why don't risk sharing considerations drive monitoring higher, as they would in the baseline APZ model? The reason is that delegation by definition splits roles: LPs own a fraction of the fund, $(1-\phi^*)\,\alpha_G^D$, but monitoring is undertaken by the GPs who own the rest: $\phi^*\alpha_G^D$.

Thus, delegation enables the separation of (LP-) ownership and (GP-) monitoring. The contract that maximizes LP welfare ensures that monitoring occurs only at the level that is privately optimal for LPs absent risk-sharing considerations. To summarize:

*Remark* 3. The delegation of blockholding breaks the link between diversification by LPs and the monitoring that occurs on their behalf. The optimal linear contract enables monitoring at a level privately optimal for LPs absent risk-sharing motives.

Having established the degree of monitoring that arises under the optimal linear contract, we now re-examine the effective holdings of LPs in order to investigate the implied risk-sharing properties. In order to do so, it is now helpful to abstract from monitoring. Imagine, counterfactually, that the LPs could trade as a single collective with endowment $\omega$ and aggregate risk tolerance $\lambda$, but could *not* monitor (directly or via their GPs), and could (by some exogenous mechanism) commit to a single round of trade, avoiding the erosion of strategic advantage that arises due to the lack of commitment. What would their optimal holdings be? In this counterfactual setting, LPs would solve the following optimization problem:

$$\max_{\alpha_{LP}} \Psi(\alpha_{LP}) - (\alpha_{LP} - \omega)P(\alpha_{LP}),$$

where $\Psi(\alpha_{LP})$is defined in (3) and $P(\alpha_{LP})$ is defined in (2), with the additional proviso that $m(\cdot) = 0$. Substituting in the expressions gives:

$$\alpha_{LP}\mu - \frac{1}{2\lambda}\alpha_{LP}^2\sigma^2 - (\alpha_{LP} - \omega)\left(\mu - \frac{1 - \alpha_{LP}}{1 - \lambda}\sigma^2\right),$$

which gives rise to $\alpha_{LP} = \lambda\frac{1+\omega}{1+\lambda}$, which is identical to the LP's effective stake under optimal delegated ownership in Corollary 1. Thus, effectively, under the optimal linear contract the LPs are able to utilize delegation to trade to their privately optimal level of risk sharing, taking into account their collective market power.

*Remark* 4. The optimal linear delegation contract enables LPs to trade to their privately optimal level of risk sharing absent monitoring motives, taking into account their collective market power.

Remarks 3 and 4 illustrate that under the optimal linear delegation contract, LPs are able to obtain both their privately optimal level of monitoring (absent risk sharing motives) as well as their privately optimal level of risk diversification (absent monitoring motives). In other words, though we started with a ostensibly specific linear contracting arrangement with a trio of parameters, $(\phi, f, \tau)$, the optimal version of such a contract fully implements the privately optimal outcome for LPs. Thus:

**Corollary 2.** $(\tau^*, \phi^*, f^*)$ *is an optimal contract.*

Finally, in order to complete our analysis of delegated blockholding and compare it to the benchmark case of APZ, we compute the overall stake held by the fund under optimal delegated ownership. We show that the optimal fund holds a smaller amount of the risky asset than the competitive risk sharing optimal allocation, which is also what a proprietary blockholder representing the same measure of traders would hold under a globally stable allocation. In particular:

**Corollary 3.** *The optimal fund holds less of the risky asset than the corresponding competitive equilibrium allocation for a trader with the same overall risk tolerance.*

## 3.1 Risk-sharing and monitoring: Delegated vs Proprietary Ownership

We are now in a position to compare our results on delegated blockholding to those of APZ's benchmark (presented in Section 2). Taking as given the existence of a proprietary trader with large risk-bearing capacity, APZ ask whether the anticipation of monitoring costs affects the degree of risk diversification in the economy. Under broad and plausible conditions, they find the answer is “no”—concentrated blockholders still trade to the competitive risk sharing allocation and monitoring occurs at that allocation. However, we show that when blockholding is achieved by optimal delegation, the picture changes dramatically, in at least two ways.

First, the delegated vehicle that is formed holds *less* of the risky asset, i.e., a smaller block, than what is implied by unconstrained risk sharing, whereas in the proprietary APZ case unconstrained risk sharing is achieved. This is because, when delegating to form a fund, the optimal contractual terms account for the fact that the fund will affect prices when trading and thus ensures that the LPs ultimately hold an amount of the risky asset that reflects their market power (and thus shades its trades downwards), as shown in Remark 4 and Corollary 3. Thus, in contrast to to the APZ proprietary benchmark, delegated blockholding results in *underdiversification* relative to the unconstrained optimum.

Second, delegation separates ownership and monitoring by allocating monitoring tasks only to a subset of participants in the fund, i.e., the GPs. The optimal delegation contract allocates an effective stake for GPs of $\omega$ (see Corollary 1) which results in a level of monitoring that would be privately optimal for LPs absent risk-sharing considerations, i.e., corresponding purely to their initial endowments (see Remark 3). As a result, the optimal delegation contract achieves strictly less monitoring than a proprietary blockholding of comparable size.

### 3.2 Microfounding contracting via Bertrand competition

Finally, we can now return to the potential microfoundation of our contracting exercise in terms of perfect competition amongst GPs. Since $\omega < \lambda$, it follows immediately that $\tau^* < \lambda$. Thus, as noted above, as long as the measure of constrained agents is not too large, we can microfound our focus on delegation contracts that maximize the payoff of LPs as being the result of Bertrand competition between GPs. A sufficient (but not necessary) condition to guarantee this is that the measure of constrained LPs is no larger than 1/3rd of the population of all agents: Then $\lambda \leq \frac{1}{2}(1-\lambda)$, and the measure of potential GPs is always more than twice the measure of optimally employed GPs: $1-\lambda > 2\tau^*$, i.e., there is always at least one further group of measure $\tau^*$ that can compete in a Bertrand manner with the optimally selected group of GPs, ensuring that LP payoffs are maximized in equilibrium.

### 3.3 Recontracting

In APZ, there is no trade-off between diversification and monitoring, because an endowment effect induces the large trader, $L$, to trade all the way to the risk-sharing optimum. Counterfactually, if any sequence of trades led to a proposed final holding level for $L$ that is strictly below her risk-sharing optimum, she would be tempted to buy a bit more because the current holding is now part of her endowment. Starting from this endowment, there will always be at least some incremental risk-sharing gains by buying a bit more, despite having to pay the full value of future monitoring in making such purchases. In the optimal contract solved above, LPs end up with a stake of $\frac{\lambda}{1+\lambda}(1+\omega)$, which is greater than their initial endowment of $\omega$, but monitoring occurs at a level corresponding to an ownership of $\omega$. Hence, a discerning reader may wonder whether a variant of the APZ endowment effect may come into play wherein the LPs now wish to recontract with GPs to reflect their new endowment. Could the possibility of such recontracting revive the APZ result, wherein the LPs achieve the risk-sharing optimum holding of $\lambda$, and monitoring occurs at a commensurate level?

In principle, the LPs as a group would indeed like to recontract with GPs to form a fund that monitors more intensely. To see this, consider a situation where the equilibrium contract from above is signed and the fund trades to the stable allocation, but then an unexpected opportunity arises to dissolve the existing fund and start a new one prior to any monitoring taking place. We then have a repeat of the model above starting from an aggregate LP endowment of $\frac{\lambda}{1+\lambda}(1+\omega)$ instead of $\omega$, which may lead to a new fund that will monitor at a level corresponding to ownership of $\frac{\lambda}{1+\lambda}(1+\omega)$.[4]

However, unlike repeated trading, which is always feasible, repeated contracting may be impossible because it is sensitive to free riding: as shown above in Proposition 2 and discussed in Remark 2, as soon as LPs have an endowment higher than $\hat{\omega} < \lambda$, it is not possible to form a fund. For such high endowments, the risk-sharing benefits to individual LPs is too small, and thus each individual LP would benefit by deviating and staying out of the fund (if it is formed), thus saving themselves the fees that must be paid to the GPs. As a result, any endowment level $\omega < \hat{\omega}$ for which a fund can be formed but $\frac{\lambda}{1+\lambda}(1+\omega) > \hat{\omega}$ holds would not be subject to recontracting. This clearly holds for some positive measure set of endowments $\Omega_S = \{\omega' > 0 : \omega' \leq \hat{\omega} < \frac{\lambda}{1+\lambda}(1+\omega') < \lambda\}$. Thus, the possibility of recontracting does not revive the APZ result.

# 4 Mutual Funds and Hedge Funds: Clientele, Fees, and Engagement

Our analysis of optimal delegation arrangements has implications for the asset management industry and the degree to which different types of asset managers engage in the monitoring of portfolio firms. Proposition 2 implies that the skin in the game of the asset managers (GPs), $\phi^* = \frac{(1+\lambda)\omega}{2\lambda\omega+\lambda+\omega}$, and their total effective investment in the risky asset, $\omega$, are both

[4]As noted in Lemma 2, GPs in the fund and unconstrained investors outside the fund hold the same effective stakes after trading, so the new fund formation problem is isomorphic to the original problem with different endowments.

increasing in the endowment of each underlying investor (LP) in the fund. In turn, the level of equilibrium monitoring undertaken by the fund, $\frac{\omega}{\gamma}$, increases in asset managers' effective stake. Thus, within the constraint under which delegated blockholding arises in equilibrium ($\omega < \hat{\omega} \in (0, \lambda)$), if fund investors have a relatively high endowment of the risky asset, they will invest in funds where managers take a larger personal stake in the fund; these funds monitor more aggressively. In contrast, if fund investors have a relatively small endowment of the risky asset, they will invest in funds where managers take a small personal stake in the fund; these funds monitor their portfolio firms very little.

The above depiction of asset management resonates with key characteristics of different types of asset management firms observed in reality. Relatively poor real world investors tend to invest in mutual funds. It is well documented that mutual fund managers invest very little in their funds: according to Khorana, Servaes, and Wedge (2007) some 57% of mutual fund managers do not invest at all in their constituent funds, and the average self-investment among the rest is 0.04%. Finally, mutual funds are notorious for being relatively muted in their engagement of portfolio firms (e.g., Bebchuck, Cohen, and Hirst 2017). In contrast, relatively wealthy individuals tend to invest in hedge funds, which typically have minimal net worth requirements. Hedge funds managers are well known to self-invest significantly in the fund (estimates in the literature range from 7% of fund assets under management in Agarwal, Daniel, and Naik 2009 to 20% in He and Krishnamurthy 2013). It is also well documented that hedge funds play a far more active role in the monitoring of their portfolio firms (Brav, Jiang, and Kim 2010).

Our results also imply that stake size may not be a good predictor of monitoring intensity. With proprietary blocks, larger stakes imply more monitoring because stake size directly determines monitoring intensity. However, with delegated blocks the fund's internal incentive structure separates monitoring incentives from stake size. As a result, funds with smaller stakes might actually monitor more intensively than those with larger stakes depending on their investor clientele. Specifically, the total delegated block size in our model is $\omega + \frac{\lambda}{1+\lambda}(1+\omega)$ which is increasing both in the number of LPs and in the aggregate endowment of the LPs. Potentially, therefore, funds with many investors holding limited endowments (large $\lambda$, small $\omega$) can have large blocks with very little monitoring. In contrast, funds with fewer investors who hold high endowments (small $\lambda$, relatively large $\omega$) can have relatively small blocks with significantly more monitoring.

As an illustrative example, compare a fund with a relatively large number ($\lambda = 15\%$) of investors with very limited endowments of the risky asset ($\omega = 0.1\%$) versus a fund with a relatively small number ($\lambda = 5\%$) of investors with relatively high endowments of the risky asset ($\omega = 0.5\%$). The former fund would hold approximately 13% of the firm, GPs would own a very small fraction—0.8%—of the fund's assets under management, and for $\gamma = 0.1$, monitoring would occur at a low intensity of $\frac{\omega}{\gamma} = 0.01$. The latter fund would hold approximately 5% of the firm, i.e., a much smaller stake; GPs would own a much larger fraction—9.5%—of the fund's assets under management, and for $\gamma = 0.1$, monitoring would occur at five-times the intensity of the other fund, $\frac{\omega}{\gamma} = 0.05$.

While our model is not ideally suited for calibration, it is noteworthy that these block sizes, GP-ownership stakes, and monitoring intensity are broadly in line with observations about mutual fund families and hedge funds. Large families like Blackrock or Fidelity often own well over 10% of US corporations but arguably do not monitor much, while their managers typically have very small stakes in the fund (as discussed above). In contrast, activist hedge funds hold a median stake of around 5-6% in target firms, monitor intensively, and—as discussed above—GPs often hold a personal stake of around 10% of the fund's assets under management. Our results are also consistent with Nockher (2022), who finds that smaller blockholders, and particularly those with a larger percentage of their fund invested in a given firm, tend to be more engaged monitors than larger blockholders.

## 5 Conclusion

Blockholder monitoring is important, but the determinants of long-term block sizes and the resulting implications for the degree of monitoring are not fully understood. The existing theoretical literature devoted to this question focuses only on proprietary blockholding, whereas modern markets are dominated by delegated asset managers. We present a simple model of delegated trading and monitoring to examine the economics of concentrated ownership and blockholder monitoring in financial markets dominated by institutional investors.

Our analysis shows that delegation has important consequences for both block sizes and monitoring. In particular, optimal delegation contracts allow for the separation of diversification and monitoring motives. This can lead to less monitoring and inferior risk sharing relative to proprietary blocks, but gives rise to monitoring and risk sharing benefits where proprietary blocks would not exist.

At an applied level, our model illustrates how some commonly observed characteristics of asset management firms—the clientele they serve, the extent of managerial self-investment, and the degree to which they monitor portfolio firms—can arise as a result of optimal contracting with fund investors. Further, our results imply that block size may not be a good predictor of monitoring intensity because the fund's internal incentive structure separates monitoring incentives from stake size. Finally, given that we conclude that active asset managers may endogenously avoid utilizing their full risk-bearing capacity to hold concentrated positions, our analysis indirectly highlights the importance of the governance role of index asset managers—who mechanically hold concentrated stakes—in corporate governance.

# Appendix

**Proof of Proposition 1:** We begin with condition (i) of the globally stable allocation. Combining definition (3) with the selected monitoring level (1) and the market clearing price (2), the optimization problem can be written as:

$$\max_{\alpha} \alpha \left(\mu + \frac{\alpha}{\gamma}\right) - \frac{1}{2}\gamma\left(\frac{\alpha}{\gamma}\right)^2 - \frac{1}{2\lambda}\alpha^2\sigma^2 - \Psi(\alpha_G) - (\alpha - \alpha_G)\left(\mu + \frac{\alpha_G}{\gamma} - \frac{1-\alpha_G}{1-\lambda}\sigma^2\right),$$

giving rise to the following first order condition:

$$\mu + \frac{2\alpha}{\gamma} - \frac{\alpha}{\gamma} - \frac{1}{\lambda}\alpha\sigma^2 - \left(\mu + \frac{\alpha_G}{\gamma} - \frac{1-\alpha_G}{1-\lambda}\sigma^2\right) = 0.$$

Now, setting $\alpha = \alpha_G$ above and solving gives:

$$\frac{1}{\lambda}\alpha_G\sigma^2 = \frac{1-\alpha_G}{1-\lambda}\sigma^2, \text{i.e., } \alpha_G = \lambda.$$

Now, we turn to condition (ii) of the globally stable allocation to verify that $\Psi(\lambda) - \Psi(\omega) - (\lambda - \omega)P(\lambda) > 0$ for all $\omega \neq \lambda$. This is equivalent to showing that $\omega = \lambda$ is a global maximum of the function

$$\Psi(\omega) - \omega P(\lambda) = \omega\left(\mu + \frac{\omega}{\gamma}\right) - \frac{1}{2}\gamma\left(\frac{\omega}{\gamma}\right)^2 - \frac{1}{2\lambda}\omega^2\sigma^2 - \omega\left(\mu + \frac{\lambda}{\gamma} - \sigma^2\right).$$

To verify this we first note that the first order condition

$$\mu + \frac{2\omega}{\gamma} - \frac{\omega}{\gamma} - \frac{1}{\lambda}\omega\sigma^2 - \left(\mu + \frac{\lambda}{\gamma} - \sigma^2\right) = 0$$

is satisfied at $\omega = \lambda$. We then evaluate the second order condition at $\omega = \lambda$: $\frac{1}{\gamma} - \frac{\sigma^2}{\lambda}$. This is strictly negative as long as $\lambda < \gamma\sigma^2$ as required.■

**Proof of Lemma 1:** We begin with condition (i) of the globally stable allocation. Combining definition (6) with the selected monitoring level (4) and the market clearing price (5), the optimization problem can be written as:

$$\max_{\alpha} \phi\alpha\left(\mu+\frac{\phi\alpha}{\gamma}\right)-\frac{1}{2}\gamma\left(\frac{\phi\alpha}{\gamma}\right)^2-\frac{\phi^2\alpha^2\sigma^2}{2\tau}-\Psi(\alpha_G^D)-\phi(\alpha-\alpha_G^D)\left(\mu+\frac{\phi\alpha_G^D}{\gamma}-\frac{1-\alpha_G^D}{1-\lambda-\tau}\sigma^2\right),$$

giving rise to the following first order condition:

$$\phi\mu+\phi^2\frac{2\alpha}{\gamma}-\phi^2\frac{\alpha}{\gamma}-\frac{1}{\tau}\phi^2\alpha\sigma^2-\phi\left(\mu+\frac{\phi\alpha_G^D}{\gamma}-\frac{1-\alpha_G^D}{1-\lambda-\tau}\sigma^2\right)=0.$$

Now, setting $\alpha=\alpha_G^D$ above and solving gives

$$\frac{1}{\tau}\phi^2\alpha_G^D\sigma^2=\phi\frac{1-\alpha_G^D}{1-\lambda-\tau}\sigma^2, \text{i.e., } \alpha_G=\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}.$$

Now, we turn to condition (ii) of the globally stable allocation to verify that $\Psi^D(\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau})-\Psi^D(\omega)-\phi(\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}-\omega)P^D(\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau})>0$ for all $\omega\neq\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}$. This is equivalent to showing that $\omega=\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}$ is a global maximum of the function $\Psi^D(\omega)-\phi\omega P^D(\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau})$, i.e.,

$$\phi\omega\left(\mu+\frac{\phi\omega}{\gamma}\right)-\frac{1}{2}\gamma\left(\frac{\phi\omega}{\gamma}\right)^2-\frac{1}{2\tau}\omega^2\phi^2\sigma^2-\phi\omega\left(\mu+\frac{\phi\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}}{\gamma}-\frac{1}{\tau/\phi+1-\lambda-\tau}\sigma^2\right).$$

To verify this we first note that the first order condition

$$\phi\mu+\frac{2\phi^2\omega}{\gamma}-\frac{\phi^2\omega}{\gamma}-\frac{1}{\tau}\omega\phi^2\sigma^2-\phi\left(\mu+\frac{\phi\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}}{\gamma}-\frac{1}{\tau/\phi+1-\lambda-\tau}\sigma^2\right)=0$$

is satisfied at $\omega = \frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}$. We then evaluate the second order condition at $\omega = \frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}$: $\frac{\phi^2}{\gamma} - \frac{\phi^2\sigma^2}{\tau}$. This is strictly negative as long as $\tau < \gamma\sigma^2$ as required.■

**Proof of Lemma 2:**The per-GP effective allocation is $\frac{\phi\alpha_G^D}{\tau} = \frac{\phi}{\tau+\phi-\phi(\lambda+\tau)}$. The unconstrained investors outside the fund hold an aggregate stake of $1-\alpha_G^D = \frac{\phi-\phi(\lambda+\tau)}{\tau+\phi-\phi(\lambda+\tau)}$, leading to a per-investor allocation of $\frac{1}{1-\lambda-\tau}\frac{\phi-\phi(\lambda+\tau)}{\tau+\phi-\phi(\lambda+\tau)} = \frac{\phi}{\tau+\phi-\phi(\lambda+\tau)}$. ■

**Proof of Proposition 2:** The fee $f(\phi, \tau)$ is set to just meet the participation constraint of individual GPs to ensure that the chosen mass $\tau$ joins the fund given the chosen skin in the game parameter $\phi$. The fund's total endowment is given by $\omega + \tau\frac{1-\omega}{1-\lambda}$ since the $\tau$ mass of GPs shares an aggregate endowment of $1-\omega$ among the entire $1-\lambda$ population of unconstrained investors. The per-GP payoff for those who join the fund is given by

$$\frac{1}{\tau}\left[\Psi^D(\alpha_G^D) - \phi\left(\alpha_G - \omega - \tau\frac{(1-\omega)}{1-\lambda}\right))P(\alpha_G^D) + \lambda f\right]. \tag{9}$$

The per-investor payoff for unconstrained investors who do not join the fund is

$$\frac{1}{1-\lambda-\tau}\left[\Psi_U^D(\alpha_G^D) - \left(1 - \alpha_G^D - \left(1 - \omega - \tau\frac{(1-\omega)}{1-\lambda}\right)\right)P^D(\alpha_G^D)\right],$$

where $\Psi_U^D(\alpha) = (1-\alpha)(\mu_0 + m^D(\alpha)) - \frac{(1-\alpha)^2\sigma^2}{2(1-\lambda-\tau)}$ is the aggregate certainty equivalent payoff of the mass of $1-\lambda-\tau$ unconstrained investors outside of the fund who hold an aggregate stake of $1-\alpha$ given that the fund holds a stake of $\alpha$. By defecting from the fund unilaterally, any given GP who is supposed to join the fund can enjoy the latter payoff. Thus, their participation constraint will be met as long as $f$ is set to make these two payoffs equivalent.

Thus, let

$$f(\phi,\tau)=\frac{1}{\lambda}\left(\begin{array}{c}\frac{\tau}{1-\lambda-\tau}\left[\Psi_U^D(\alpha_G^D)-\left(1-\alpha_G^D-\left(1-\omega-\tau\frac{(1-\omega)}{1-\lambda}\right)\right)P^D(\alpha_G^D)\right]\\ -\left[\Psi^D(\alpha_G^D)-\phi(\alpha_G-\omega-\tau\frac{(1-\omega)}{1-\lambda})P^D(\alpha_G^D)\right]\end{array}\right) \tag{10}$$

To determine the optimal $\phi$ and $\tau$, we substitute in the expression for $f(\phi,\tau)$ from (10) and $\alpha_G^D=\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}$ then take the $\phi$-derivative of the per-LP payoff

$$\begin{gathered}\frac{1}{\lambda}\alpha_G^D(1-\phi)\left(\mu+\frac{\alpha_G^D\phi}{\gamma}\right)+\frac{1}{\lambda}(1-\phi)\left(\alpha_G^D-\omega-\frac{(1-\omega)\tau}{1-\lambda}\right)\left(\mu+\frac{\alpha_G^D\phi}{\gamma}-\frac{(1-\alpha_G^D)\sigma^2}{1-\lambda-\tau}\right)\\ -\frac{\left(\alpha_G^D\right)^2\sigma^2(1-\phi)^2}{2\lambda^2}-f(\phi,\tau)\end{gathered} \tag{11}$$

to obtain:

$$\frac{\tau\left(\tau\left((\lambda-1)\lambda\omega\phi-\gamma\sigma^2(\phi-1)(\lambda\omega-1)\right)+(\lambda-1)\lambda\gamma(\omega+1)\sigma^2(-\phi)+\lambda\tau^2(\omega(\phi-1)+\phi)\right)}{\lambda^2\gamma(\phi(\lambda+\tau-1)-\tau)^3}. \tag{12}$$

Now, the first order condition with respect to $\phi$ gives a unique local optimum:

$$\phi(\tau)=\frac{\tau\left(\gamma\sigma^2(1-\lambda\omega)+\lambda\omega\tau\right)}{\gamma\sigma^2\tau(1-\lambda\omega)-(\lambda-1)\lambda\gamma(\omega+1)\sigma^2+\lambda(\omega+1)\tau^2+(\lambda-1)\lambda\omega\tau}. \tag{13}$$

Taking the second derivative of (12) and evaluating it at (13) gives

$$\frac{\left(\gamma\sigma^2\tau(\lambda\omega-1)+(\lambda-1)\lambda\gamma(\omega+1)\sigma^2-\lambda(\omega+1)\tau^2-(\lambda-1)\lambda\omega\tau\right)^4}{\lambda^2\gamma\tau^2\left(\left(\lambda^2-1\right)\gamma\sigma^2-\lambda\tau^2\right)^3},$$

which is negative, confirming a unique maximum. Inserting (13) and (10) into (11) and simplifying gives:

$$\frac{\gamma^2\sigma^2\left(2\left(\lambda^2-1\right)\omega\mu-\sigma^2\left(\lambda\omega^2+\lambda-2\omega\right)\right)+\tau^2\left(\gamma\sigma^2-\lambda\omega(2\gamma\mu+\omega)\right)+2\gamma\omega\sigma^2\tau(\lambda\omega-1)}{2\lambda\gamma\left(\left(\lambda^2-1\right)\gamma\sigma^2-\lambda\tau^2\right)}.$$

To solve for the optimal $\tau$, we take the $\tau$-derivative to obtain:

$$\frac{\sigma^2\left(\omega\left(\lambda^2-\lambda\tau-1\right)+\tau\right)\left(\gamma\sigma^2(\lambda\omega-1)-\lambda\omega\tau\right)}{\lambda\left(\left(\lambda^2-1\right)\gamma\sigma^2-\alpha\tau^2\right)^2}.$$

The first order condition gives two potential values for the optimal $\tau$:

$$\tau\in\left\{\frac{\left(1-\lambda^2\right)\omega}{1-\lambda\omega},\frac{\gamma\sigma^2(\lambda\omega-1)}{\lambda\omega}\right\}$$

of which only the former is positive. We check the second order condition and find that it is indeed negative at $\tau=\frac{\left(1-\lambda^2\right)\omega}{1-\lambda\omega}$, verifying that this is an unique positive maximand. Thus, $\tau^*=\frac{\left(1-\lambda^2\right)\omega}{1-\lambda\omega}$, and substituting it into (13) gives $\omega^*=\frac{(1+\lambda)\omega}{2\lambda\omega+\lambda+\omega}$. Finally, substituting $\tau^*$ and $\phi^*$ into (10) gives (8).

Finally, we must check that it is incentive compatible for each LP to remain inside the fund instead of defecting from it. First, by inserting the optimal $(\phi^*, f^*, \tau^*)$ into (11) we obtain the equilibrium payoff of each LP as follows:

$$\left(\frac{\sigma^2\left(\lambda\omega^2+\lambda-2\omega\right)}{1-\lambda^2}+\frac{\omega^2}{\gamma}+2\omega\mu\right)/2\lambda \tag{14}$$

Each LP has the option to defect from the fund, consume their own endowment of $\omega/\lambda$, enjoy the benefits of monitoring by the fund, but pay no fees. Algebraic computations show that, at the optimal contracting parameters $(\phi^*, f^*, \tau^*)$ and when the fund holds the globally stable allocation $\alpha_G^D$, such a deviation payoff for each LP (with unit risk tolerance) can be expressed as follows:

$$\frac{\omega\left(2\lambda(\gamma\mu+\omega)-\gamma\omega\sigma^2\right)}{2\lambda^2\gamma}. \tag{15}$$

Subtracting (15) from (14) gives

$$\frac{\frac{\sigma^2(\lambda-\omega)^2}{1-\lambda^2}-\frac{\lambda\omega^2}{\gamma}}{2\lambda^2}$$

which is clearly positive for $\omega = 0$, clearly negative for $\omega = \lambda$, and clearly decreasing in $\omega$ for $\omega \in (0, \lambda)$. Thus, by continuity, there clearly exists a $\hat{\omega} \in (0, \lambda)$ such that for $\omega \leq \hat{\omega}$, no LP will wish to deviate and leave the fund.■

**Proof of Corollary 1:** Plugging the expressions for $\phi^*$ and $\tau^*$ from Proposition 2 into (7) yields $\omega + \frac{\lambda}{1+\lambda}(1+\omega)$. Multiplying this with the expression for $\phi^*$ yields $\omega$. ■

**Proof of Corollary 3:** The fund holds $\alpha_G^D = \frac{\tau^*/\phi^*}{\tau^*/\phi^*+1-\lambda-\tau^*}$ of the risky asset in equilibrium. The fund is made up of agents of measure $\lambda + \tau^*$ and thus the collective risk tolerance of this group of agents is $\lambda + \tau^*$. In a competitive equilibrium, such a collective of agents will hold $\lambda + \tau^*$ of the risky asset. Using the expressions in Proposition 2 we have:

$$\frac{\tau^*}{\phi^*} = \frac{\left(1-\lambda^2\right)\omega}{1-\lambda\omega}\frac{2\lambda\omega+\lambda+\omega}{(1+\lambda)\omega} = \frac{(1-\lambda)\left(2\lambda\omega+\lambda+\omega\right)}{1-\lambda\omega}.$$

We first show that $\tau^*/\phi^* < \lambda + \tau^*$. Assume the contrary. This implies that:

$$\frac{(1-\lambda)\left(2\lambda\omega+\lambda+\omega\right)}{1-\lambda\omega} \geq \lambda + \frac{\left(1-\lambda^2\right)\omega}{1-\lambda\omega},$$

which simplifies to $\lambda\left(\omega-\lambda\right) \geq 0$, which is a contradiction because $\lambda > 0$ and $\omega \leq \lambda$. Having now shown that $\frac{\tau^*}{\phi^*} < \lambda + \tau^*$, we now observe that $\alpha_G^D = \frac{\tau^*/\phi^*}{\tau^*/\phi^*+1-\lambda-\tau^*} < \frac{\lambda+\tau^*}{\lambda+\tau^*+1-\lambda-\tau^*} = \lambda + \tau^*$.■

# Bibliography

Admati, A., P. Pfleiderer, and J. Zechner 1994. Large shareholder activism, risk sharing, and financial market equilibrium. Journal of Political Economy 102, 1097-1130.

Agarwal, Vikas, Naveen D. Daniel, and Narayan Y. Naik, 2009, Role of managerial incentives and discretion in hedge fund performance, Journal of Finance 64, 2221–2256.

Back, K., P. Collin-Dufresne, V. Fos, T. Li, and A. Ljunqvist 2018. Activism, strategic trading, and liquidity. Econometrica. 86, 1431–1463.

Bebchuk, L. A., A. Cohen, and S. Hirst 2017. The agency problems of institutional investors. Journal of Economic Perspectives. 31, 89–102.

Bhattacharya, S. and P. Pfleiderer 1985. Delegated Portfolio Management. Journal of Economic Theory 36, 1-25.

Bolton, P. and E. von Thadden 1998. Blocks, Liquidity, and Corporate Control. Journal of Finance 53, 1-25.

Brav, A., W. Jiang, and H. Kim 2010. Hedge fund activism: A review. Foundations and Trends in Finance. 4, 185–246.

Brav, A., A. Malenko, and N. Malenko 2022. Corporate governance implications of the growth in indexing. ECGI working paper 849/2022.

Dasgupta A, V. Fos, Z. Sautner 2021. Institutional Investors and Corporate Governance. Foundations and Trends in Finance, 12, 276–394.

Dasgupta, A. and G. Piacentino 2015. The wall street walk when blockholders compete for flows. Journal of Finance 70, 2853– 2896.

DeMarzo, P. and B. Urosevic 2006. Ownership Dynamics and Asset Pricing with a Large Shareholder. Journal of Political Economy 114, 774-815.

Diamond, D. and P. Dybvig 1983. Bank Runs, Deposit Insurance, and Liquidity. Journal of Political Economy 91, 401–419.

Faure-Grimaud, A. and D. Gromb 2004. Public trading and private incentives. Review of Financial Studies 17, 985–1014.

Grossman, Sanford J., and Oliver D. Hart, 1980, Takeover bids, the free-rider problem, and the theory of the corporation, Bell Journal of Economics 11, 42–64.

He, Zhiguo, and Arvind Krishnamurthy, 2013, Intermediary asset pricing. American Economic Review 103, 732–770.

Kahn, C. and A. Winton 1998. Ownership structure, speculation, and shareholder intervention. Journal of Finance 53, 99–129.

Khorana, Ajay, Henri Servaes, and Lei Wedge, 2007, Portfolio manager ownership and fund performance. Journal of Financial Economics 85, 179–204.

Kyle, A. and J.-L.Vila 1991. Noise trading and takeovers. RAND Journal of Economics. 22, 54–71.

Maug, E. 1998. Large shareholders as monitors: Is there a trade-off between liquidity and control? Journal of Finance 53, 65-98.

Nockher, F. 2022. The Value of Undiversified Shareholder Engagement. Working paper, University of Pennsylvania.

Shleifer. A. and R. Vishny 1986. Large shareholders and corporate control. Journal of Political Economy 94, 461–488.